BIOINFORMATICS IN THE BIOTECHNOLOGY INSTITUTES OF NORTHERN INDIA : A SURVEY AMONG SCIENTISTS

DISSERTATION

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE OF THE

**Master
of
Library & Information Science**

(2001-2002)

By

IRSHAD AHMAD
Roll No. 2k1-LSM-408
Enrol No. X-6638

Under the Supervision of

MR. NAUSHAD ALI P.M.
(LECTURER)

DEPARTMENT OF LIBRARY & INFORMATION SCIENCE
ALIGARH MUSLIM UNIVERSITY
ALIGARH (INDIA)

2002

ALIGARH MUSLIM UNIVERSITY, ALIGARH

DEPARTMENT OF LIBRARY
AND
INFORMATION SCIENCE

Phone { EPBX : 700916, 20-23, 26 Ext 193/4
Direct : 700039
Fax : 91-0571-470528, 471221

Ref. No..........................
Dated 28 AUG 2002

Certificate

This is to certify that ***Mr. Irshad Ahmad*** has completed his dissertation entitled ***"BIOINFORMATICS IN THE BIOTECHNOLOGY INSTITUTES OF NORTHERN INDIA: A SURVEY AMONG SCIENTISTS"***, in partial fulfilment of the requirements for the degree of **Master of Library & Information Science** (2001-2002). He has conducted the work under my supervision and guidance. I deem it fit for submission.

Naushad Ali P.M.
Lecturer

Dedicated
To My
Loving Parents
&
Other
Family Members

CONTENTS

ACKNOWLEDGEMENT

Praise be the Almighty **ALLAH**, the most merciful and most beneficent, who showed me the path of righteousness and blessed me with strength to fulfil the requirements of M.L. & I.Sc. degree.

Besides, it is to my teacher and supervisor ***Mr. Naushad Ali P.M.*** Lecturer, Department of Library & Information Science, A.M.U., that to whom I am highly indebted, for his help and guidance without which the present work could not have seen the light of day.

I wish to express my deep sense of gratitude to ***Mr. Syed Mustafa Kazim Qaiser Zaidi***, Reader and Chairman, Department of Library & Information Science, A.M.U. for lighting the path of optimisation.

I want to express my deep sense of gratitude to ***Prof. Shabahat Husain***, Ex-Chairman, Department of Library & Information Science, A.M.U. for his valuable suggestions during my course of study.

I would like to pay my grateful and sincere thanks to ***Mr. M. Masoom Raza, Miss Sudharma Haridasan, Miss Nishat***

Fatima* and *Mr. Asif Fareed Siddiqui, Lecturers, Department of Library & Information Science, A.M.U. for their help and encouragement from time to time.

I am also thankful to all members of the non-teaching staff of the Department of Library & Information Science, A.M.U. especially ***Mr. Riaz Abbas, Mr. Khawaja Moin*** and our great ***Babu Bhai***.

My heartful thanks are due to all my friends, classmates and well wishers.

I will be failing in my duty, if I do not express my sincere gratitude and indebtedness to my parents and other members of my family, who have been inspiring me and giving immense moral support during various stages of the study.

Irshad Ahmad

(IRSHAD AHMAD)

LIST OF ABBREVIATIONS/ ACRONYMS

1.	A°	Angstrom
2.	BLAST	Basic Local Alignment Tools
3.	CAS	Chemical Abstracts Service
4.	CAS	Current Awareness Service
5.	CDFD	Centre for DNA Fingerprinting and Diagnostic
6.	CD-ROM	Compact Disk-Read Only Memory
7.	CIFA	Central Institute of Fresh-water Aquaculture
8.	DDBJ	Databank of Japan
9.	DICS	Distributed Information Centres
10.	D.N.A.	Deoxyribonucleic Acid
11.	DOE	Department of Energy
12.	E-mail	Electronic-mail
13.	EMBL	European Molecular Biology
14.	HGO	Human Genome Organisation
15.	HGP	Human Genome Project
16.	MIT	Massachusetts Institute of Technology
17.	N.C.B.I.	National Center for Biotechnology Informations
18.	PDB	Protein Data Bank
19.	PIR	Protein Information Resource
20.	R & D	Research and Development
21.	R.N.A.	Ribonucleic Acid
22.	SDI	Selective Dissemination of Information
23.	TCP/IP	Transmission Control Protocol/Internet Protocol

LIST OF TABLES

CHAPTER-I

INTRODUCTION

BIOINFORMATICS

INTRODUCTION:

Over the past few decades, major advances in the field of molecular biology, coupled with advances in genomic technologies, have led to an explosive growth in the biological information generated by the scientific community. The deluge of genomic information has, in turn, led an absolute requirement for computerized databases to store, organize and index the data and for specialized tools to view and analyse the data.

According to Bouski, information science and technology (informatics) became a serious issue for biologists in the mid 1970s following the development of rapid DNA sequencing techniques. Since then, the amount of sequence data (and also gene mapping and protein crystal structure data) has grown exponentially and is now also being fuelled by data emerging from the world-wide Human Genome Mapping Project. The growth rate of Genbank and EMBL (European Molecular Biology Laboratory) databases has been exponential for the last five years; the latest release of Genbank (release 80.0) contains 164 megabases of sequence and the size currently doubling every twenty-one months. It is expected that over the next decade, biomolecular databanks will grow between seven and sixty-fold.

What is Bioinformatics?

Bioinformatics is the field of science in which biology, computer science, and information technology merge to form a single discipline. The ultimate goal of the field is to enable the discovery of new biological insights as well as to create a global perspective from which unifying principles in biology can be discovered. At the beginning of the "genomic revolution", a bioinformatics concern was the creation and maintenance of a database to store biological information such as nucleotide and aminoacid sequences.

Information is a powerful tool for research and development activities. It is also useful for the development of biologists, bioinformaticians, and researchers in the field of biology. Dissemination of information is a prerequisite for any development scheme meant for bringing change in all walks of life. Thus, bioinformatics has a significant role in the developmental issues of concerned fields.

The word biology consists of two words i.e., botany and zoology which mean the study of living cells. The cell is fundamental unit of living things. The study of cells is called cytology. The cytosine, a nitrogenous base derived from pyrimidine, is one of several types of bases that are incorporated into a nucleic acid molecule. Nucleic acids are composed of a five carbon sugar bound to a phosphoric acid, along with a nitrogen base. Deoxyribonucleic acid (D.N.A.), the hereditary material of most living organisms, consists of the five sugar dioxyribose with a phosphate linkage, to which is attached cytosine or any of three other bases, which together form two complementary pairs. Cytosine's complementary base in the DNA molecule is guanine.

An organism's hereditary and functional information is stored as DNA, RNA and proteins, all are linear chains composed of smaller molecules. DNA is made up of four deoxyribonucleotidis (adenine, thrymine, cytosine and guanine). RNA is made up from the four ribonucleotides (adenine, uracil, cytosine and guanine) and proteins are made up from twenty amino acids.

Now, it is quite obvious that acquisition, collection, storage, and retrieval of such a huge date and facts of biology and biotechnology are very difficult. But with the help of biotechnology and Information technology i.e. Bioinformatics it is now possible to some extent. The basic difference between biotechnology and Bioinformatics is as follows:

The word Biotechnology is made up from biology and technology. Thus biotechnology is a technology that uses living organisms, or components of living organisms or biological process to make or modify products and to improve plants, animals or micro organism for specific use/service of mankind.

Similarly, Bioinformatics is derived from two words i.e. Biotechnology and Information Technology. Thus, Bioinformatics is a new term referring to the discipline that employs computers to store, retrieve, analyse and assist in understanding biological information. It is a newly emerging interdisciplinary research area spanning a range of specialities that include biology, biophysics, computer science, mathematics and statistics.

Bioinformatics is defined as a scientific discipline that compasses all the aspects of biological information like acquisition, processing, storage, analysis, distribution and interpretation that combines the tools and techniques of mathematics, computer science and biology with the aim of understanding the biological significance of a variety of data.

In other words Bioinformatics is the subset of the larger field of computational biology, the application of quantitative analytical technologies in modelling biological systems. The science of informatics is concerned with the representation, manipulation, distribution, maintenance and use of information, particularly in digital form. Basically, bioinformatics is the combination of two words i.e. biology and informatics. The functional aspect of bioinformatics is the representation, storage and distribution of data. Intelligent design of data formats and databases, creation of tools to query these databases and development of user interfaces that bring together different tools to allow the user to ask complex questions about the data are all aspects of the development of bioinformatics infrastructure. Developing analytical tools to discover knowledge in data is the second aspect of bioinformatics.

According to a practical guide book the bioinformatics means the Analysis of Genes and proteins makes computational biology accessible to scientists at all levels of expertise, including those with no formal computer training. It cuts through the overwhelming array of existing tools and databases helping the reader design and implement a successful sequence analysis strategy. Presented by leading anthorities in computational biology, this edited volume covers gamnt of topics, from using software and internet resources to submitting DNA sequences to databases.

According to Boguski the term bioinformatics to be wide in scope involving computational analysis, databases and 'everything from laboratory automation and data acquisition to electronic publishing.'

According to Andrewlyall of Glaxo (personal communication) says the Glaxo interpretation of the word is made up of three elements as follows:

1. Computational genetics-encompassing activities such as the Human Genome Mapping Project and covering physical and logical genetic mapping.
2. Computation relating to molecular genetics, including sequence determination. This would also cover the automatic reading of data from automatic gene – sequencing machines.
3. Computation relating to three dimensional proteins structure determination.

Both of these definitions illustrate that the word Bioinformatics can be interpreted to describe virtually any information – related activity applied to the sciences of genetics or molecular biology.

According to whilst, bioinformatics means to include the storage, retrieval and analysis of nucleic acid and aminoacid sequence but will exclude three dimensional structure computation, automated laboratory procedures or any thing relating to the Human Genome Project.

Why is Bioinformatics so Important?

The rationale for applying computational approaches to facilitate the understanding of various biological processes includes:

- A more global perspective in experimental design, and
- The ability to capitalize on the emerging technology of database mining – the process by which testable hypotheses as generated regarding the function or structure of a gene or protein of interest by identifying similar sequences in better characterized organisms.

Evolutionary Biology:

New insight into the molecular basis of a disease may come from investigating the function of homologs of a disease gene in model organisms. In this case, homology refers to the fact that two genes share a common evolutionary history scientists also are the term homology, or homologous, to simply mean similar, regardless of the evolutionary relationship.

Equally exciting is the potential for uncovering evolutionary relationships and patterns between different forms of life. With the aid of nucleotide and protein sequences, it should be possible to find the ancestral ties between different organisms. So far, experience, has taught us that closely related organisms have similar sequences and that more distantly related organisms have more dissimilar sequences. Proteins that show a significant sequence conservation indicating a clear evolutionary relationship are said to be form the same protein family. By studying protein folds (distinct protein building blocks) and families, scientists are able to reconstruct the evolutionary relationship between two species and to estimate the time of divergence between two organisms since they last shared a common ancestor.

Protein Modelling:

The process of evolution has resulted in the production of DNA sequences that encode proteins with specific functions. In the absence of a protein structure that has been determined by X-ray crystallography, researchers can try to predict the three – dimensional structure using protein or molecular modelling. This method uses experimentally determined protein structures (templates) to predict the structure of another protein that has a similar amino acid sequence (target).

Although molecular modelling may not be as accurate at determining a protein's structure as experimental methods, it is still extremely helpful in proposing and testing various biological hypotheses. Molecular modelling also provides a starting point for researchers wishing to confirm a structure through X-ray crystallography. As the different genome projects are producing more sequences, and because novel protein folds and families are being determined, protein modelling will become an increasingly important tool for scientists working to understand normal and disease – related processes in living organisms.

The Four Steps of Protein Modelling:

- Identify the proteins with known three – dimensional structures that are related to the target sequence.
- Align the related three – dimensional structures with the target sequence and determine those structures that will be used as templates.
- Construct a model for the target sequence based on its alignment with the template structures.
- Evaluate the model against a variety of criteria to determine if it is satisfactory.

Genome Mapping:

Genomic maps serve as a scaffold for orienting sequence information. A few years ago, a researcher wishing to localize a gene, or nucleotide sequence, was forced to manually map the genomic region of interest, a time- consuming and often painstaking process. Today, thanks to new technologies and the influx of sequence data, a number of high quality, genome – wide maps are available to the scientific community for use in their research.

Computerized maps make gene hunting faster, cheaper and more practical for almost any scientist. In a nutshell, a scientist would first use a genetic map to assign a gene to a relatively small area of a chromosome. They would then use a physical map to examine the region of interest close up, in order to determine a gene's precise location. In light of these advances, a researcher's burden has shifted from mapping a genome or genomic region of interest, to navigating a vast number of Web sites and databases.

Human Resource Development:

A number of workshops and training programmes were conducted under the BTIS programme on the use of computers and databanks in modern biology and biotechnology. Considering the importance of the subject, some institutions and University departments have introduced a formal cause of training in bioinformatics as a 3-Credit, one- semester course in their existing post graduate programmes in biotechnology DBT has initiated a long-term academic course in Bioinformatics leading to the reward of an advance diploma in bioinformatics. The course is being conducted by Madurai kamaraj University; Pune University; Jawaharlal Nehru University and Kolkatta University. There has been a sudden spurt in the demand of such professionals by small size gene-hunting companies. Many of the larger pharmaceutical companies are now seeing real value in gene mapping and sequence data and have started attracting experts from academia. In the light of these developments, the efforts of DBT are likely to be rewarding towards generation of employment opportunities. Apex Bioinformatics centre has brought out a national training calendar for the year 1998-99 of the training programmes conducted by various centres and sub-centres and will continue to publish annually calendar every year so that the training facilities are utilised by a large number of scientists and research personnel.

Human Genome Project:

A human body cell has its genes packaged in 23 pairs of chromosomes in two sets, one coming from the mother and the other from the father at the time of fertilization of the egg by the sperm. A set of these chromosomes (hence, of genes) constitutes an individual's genome. These two sets of chromosomes are homologous (of same origin) and are, therefore, similar in shape and size. They also bear similar genes in same order. Infact, since the number of chromosomes in a species, for example man (*Homo sapiens*), is fixed, so is the number of genes and their sequence in a chromosome. Therefore, with the availability of improved recombinant DNA, gene cloning, and DNA sequencing technologies by early 1980s, it was exciting to think of knowing about the nature and finer details of ones own genes, including the possibility of sequencing the estimated 3 billion nucleotide pairs that could open up great possibilities.

I. Structure of DNA:

The perfect structure of DNA was first deduced by J.D. Watson and F.H.C. Crick in 1953. Their double-helix model of DNA structure was based on two major kinds of evidence.

1. When the composition of DNA from many different organisms was analysed by E. Chargaff and Colleagues, it was observed that the concentration of thymine was always equal to the concentration of adenine and the concentration of cytosine was always equal to the concentration of guanine. This strongly suggested that thymine and adenine as well as cytosine and guanine were present in DNA with some fixed interrelationship. Ofcourse, it also necessitated that the total concentration of pyrimidines (thymine plus cytosine) always equal the total concentration of

purines (adenine plus guanine). However, the (thymine+adenine)/ (cytosine+guanine) ratio was found to vary widely in DNAs of different species e.g. viruses.

2. When X-rays are focused through isolated macromolecule or crystals of purified molecules, the X-rays are deflected by the atoms of the molecules in specific patterns, called diffraction patterns, which provide information about the organization of the components of the molecules.

These X-ray diffraction patterns can be recorded on X-ray-sensitive film just as one photographs patterns of light with a camera and light-sensitive film. Watson and Crick had available X-ray crystallographic data on DNA structure from the studies of M.H.F. Wilkins, R. Frnaklin, and their co-workers. These data indicated that DNA was a highly ordered, multiple-stranded structure with repeating sub-structures spaced every 3.4 angstroms [1 angstrom (A^0) = 10^{-8} cm] along the axis of the molecule.

On the basis of Chargaff's chemical data, Wilkins and Franklin's X-ray diffraction data, and inferences drawn from model building, Watson and Crick proposed that DNA exists as a double helix in which the two polynucleotide chains are coiled about one another in a spiral (Figure-1). Diagram (left) and space-filling model (right) of the Watson-Crick double-helix mode of the structure of DNA. In the figure A, T, G, and C represent adenine, thymine, guanine, and cytosine, respectively and P represent sugar (2-deoxyribose) and phosphate groups. (The space-filling model is based on a diagrams by M. Feughelman et al., Nature 175: 834, 1955.

The discussions at the department of Energy (DOE), USA, and among leading molecular biologists which resulted in the launching of the Human Genome Project (HGP) in 1988, with the Nobel laureate, Dr. James D. Watson being its first Director. The Project envisaged (i) to map all human genes, (ii) to construct a physical map of the entire human genome, and (iii) to determine the nucleotide sequence of all the 23 pair of human chromosomes.

Because of some dispute, Dr. Watson resigned as the head of this project in April 1992, and Francis Collins of the University of Michigan, USA, was appointed as his replacement.

It was an ambitious project which involved mapping of an estimated 1,00,000 genes and sequencing 3.2 billion nucleotide base pairs. The scientists soon realized that the project should better be a coordinated international effort. As a result International Human Genome Organization (HGO) was organized that involved a consortium of several institutions from all over the world, especially USA, Germany, France, Japan, England and Switzerland.

STRUCTURE OF DNA

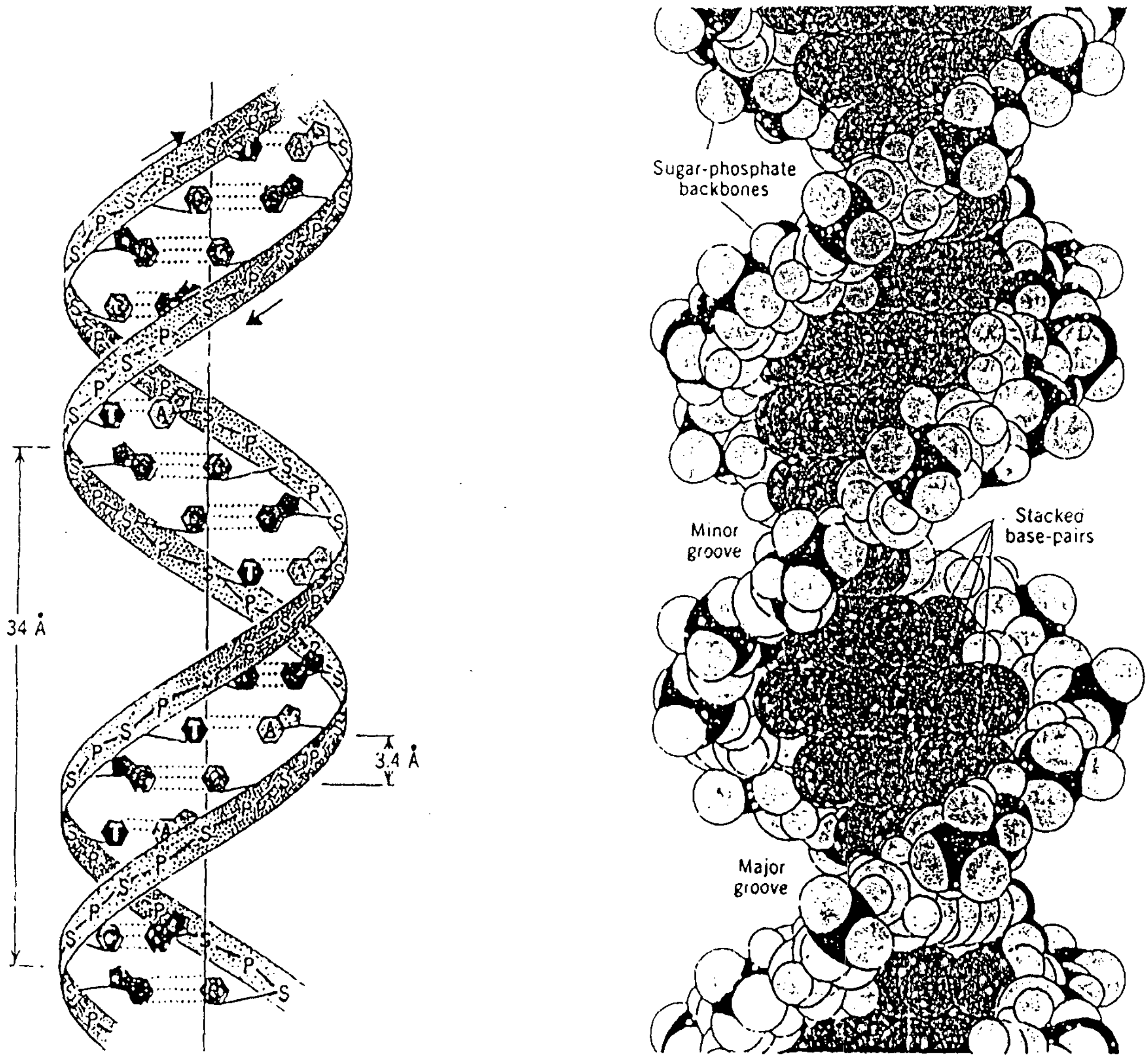

Figure-1: Diagram (left) and space-filling model (right) of the Warson-Crick double-helix of the structure of DNA.

The recent sequencing of chromosome 21 (May 2000) has identified only 225 genes compared to 545 on chromosome 22, indicating that chromosome 21 is relatively gene poor, obviously having more junk DNA (that does not code for anything).

The public – funded HGO Director, Francis Collins, announced the compilation of a 'Working draft' of the human genome and Craig Venter announced his 'first assembly' of a complete genome sequence.

Recently (February 2001) both groups have announced their draft sequence of the human genome indicating that it may contain fewer genes than anticipated. Based on the approximate ratio of the size of a typical gene (~ 3×10^4 bp) to the size of the human genome (3×10^9 bp), it was earlier estimated that there might be about 1,00,000 genes in humans. However, the recent work has indicated that their number may, in fact, be 30,000 - 35,000 – only about twice as many as in fruit fly (Drosophila). This is so because the repeat (non-coding) sequences account for more than 50 per cent of the genome, whereas the coding sequences comprise less than 5 per cent.

The studies have also indicated that the full set of proteins (proteome) encoded by the human genome is much more complex because of the complexity of the genes.

More recently (August 2001), Dr. Michael P Cooke and his Colleagues as the Genomics Institute of the Novartis Research Foundation in San Diego have compared the two sets of human genes from Celera Genomics and the International Consortium, finding that 15,000 genes are common to both. Based on their won analysis, these biologists believe that the total number of genes in the human genome should be 40,000.

Dr. Christopher Burge of the Massachusetts Institute of Technology (MIT) has also suggested in an article in the Journal Genomics that the total number of human genes is around 40,000.

However, it will perhaps take accurate years more before we have an accurate count of our genes. But, when it happens, it will pave the way for immense potentialities. With complete knowledge of the genome, it will be possible to know why we are like we are and why one suffers from a certain inadequacy. Also, it would be possible to cure the various metabolic disorders through introduction of normal functional genes (gene therapy) in place of the non-functional. However, the immense potentialities will also be prove to misuse by unscrupulous elements.

What is a Biological Database?

A biological database is a large, organized body of persistent data, usually associated with computerized software designed to update, query, and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information. For example a record associated with a nucleotide sequence database typically contains information such as contact name; the input sequence with a description of the type of molecule; the scientific name of the source organism from which it was isolated; and, often, literature citations associated with the sequence.

For researcher to benefit from the data stored in a database, two additional requirements must be met;

- Easy access to the information; and
- A method for extracting only that information needed to answer a specific biological question.

The data in GenBank is made available in a variety of ways, each tailored to a particular use, such as data submission or sequence searching.

At N.C.B.I., many of our databases are linked through a unique search and retrieval system, called Entrez. Entrez (pronounced ahn'tray) allows a user to not only access and retrieve specific information from a single database, but to access integrated information from many NCBI databases. For example, the Entrez protein database is cross linked to the Entrez taxonomy database. This allows a researcher to find taxonomic information – taxonomy is a division of the natural sciences that deals with the classification of animals and plants – for the species from which a protein sequence was derived.

Databases, it would be inappropriate to attempt to give a comprehensive review of the publicaly available molecular sequence and structure databanks. GenBank was established in 1982 at the Los Alamos National Laboratory and contains nucleic acid sequences derived from the published literature. The database also contains bibliographic data, CAS (Chemical Abstracts Service) Registry numbers and other data such as the sequence length and source organism. Whereas GenBank has a us bias, the EMBL Nucleotide Sequence Databank [8] provides a similar service to Europe and the structure of the database is very similar. In fact, GenBank and EMBL share all of their data with each other and with the DNA Databank of Japan (DDBJ) so in effect then three databases are one and the same (albeit with some time differences with respect to updates) and comprise the most comprehensive collection of nucleotide sequences. For amino acid sequences, Swiss Prot (established in 1986) is a key source. This is also a collaboration, between the Department of Medical Biochemistry at the University of Geneva and EMBL. Swiss Prot data come from the Protein Information Resource (PIR) database [10], from the translation (Via the genetic Code, from nuclic acid sequence to protein sequence) of entires from the EMBL database and directly from the literature. Entries consist of the 'Core' (primary) sequence, literature citations, taxonomic data and annotation data (protein function, secondary structure information, diseases associated with the protein, etc.). The Protein Data Bank (PDB), maintained by Brookhaven National Laboratory (Long Island, New York, USA), Contains all publicly available solved 3-D protein structures. Data include atomic co-ordinates and other data relating to how the structure was elucidated (e.g. crystallographic and NMR data). CAS has registered bio-sequences from the journal literature since 1957 although, until 1990, they were stored in electronic 'Connection tables' which define a molecule in terms of the connectivity between individual atoms and therefore could only be searched by chemical sub-structure. The Protein sequence data were enhanced in 1990 with the computer generation of amino acid sequences for all of its (approximately 150,000) protein structures; thus proteins and peptides were additionally searchable using the common short hand amino acid abbreviations. An 'exact' protein/ peptide search retrives only exact matches to the sequence query, whereas a 'subsequence' search looks for a string of amino acids anywhere in a chain (analogous to sub-structure chemical searching). Sub-sequence or exact 'family' searching is also possible, whereby each amino acid residue in a query is matched to any of its functional family members in the file structure, i.e., those that have similarities with respect to their acidity, hydrophobicity or aromaticity. The system also caters for uncommon amino acids and multi-chain systems. Since 1992, the CAS Registry file has also included nucleic acids from GenBank, as well as making the entire GenBank file available on the STN host.

Searching databases for sequence similarity (homology). The commonest questions about a given sequence databases are 'has this sequence been described in the literature before? and are there any other known sequences that are similar to my own sequence (and how similar are they?)'. Homology searches involve the use of computer program which use algorithms to calculate a similarity score between two different stretches of DNA, which is at least partly based on the summation of the number of matching nucleotide pairs within a defined local region of the complete sequence. 'Hit' sequences can therefore be ranked, exact matches having a maximum score. Many algorithms have been designed (e.g. BLAST – Basic Local Alignment Tool) and they all differ with respect to their computational speed and their sensitivity. There has been a growing need in recent years for an integrated approach towards gaining access to actual sequence data in a cross-references in the literature. Entrez, developed at the NCBI, provides this capability and contains sequence records from variety of database sources, including Gen Bank, EMBL, DDBJ, DIR, Swiss Prot, and the PDB. The sequence records are linked to the relevant literature citations from the sequence – associated subset of Medline. The retrieval software and databases are distributed on CD-ROM or as a free Internet service (Network Entrez). In addition to the 'Core' databases mentioned above, there are a great number of specialised databases, such as those dealing with a particular chromosome in the human genome, type of cell receptor or vectors.

Bioinformatics Network:

Internet is a network of computers and it is used for communication, information access, information dissemination. Internet is a global collect6ion of interconnected interlinked computer networks or network of networks. Thousands of computer networks and millions of individual's computers throughout the world are connected using Transmission Control Protocol/ Internet

Protocol (TCP/IP). The importance of information technology has been recognised for pursuing advanced research in modern biology and biotechnology. Biotechnology Information System is a network. This network consists of several information centres, which are interconnected and interlinked through computers. A bioinformatics programme, envisaged as a distributed database and network organisation, was launched during 1986-87.

This network is playing a significant role in transfer and exchange of information, scientific knowledge, technology packages, and references in the country involving scientific personnel. The entire network has emerged as a very sophisticated scientific infrastructure for bioinformatics involving state of the art computational and communication facilities. The computer communication network, linking all the bioinformatics centres, is playing a vital role in the success of the bioinformatics programme. Database development, R and D activities in bioinformatics, human resource development and a variety of services in support of Biotechnology R and D programmes and projects, have made this programme very popular and useful to the scientific community.

Access to Biotechnology Information System Net:

This is the first major S and T network in biotechnology in the country, networking 55 bioinformatics centres (an apex centre, 10 DICS and 44 sub-DICs) through satellite and terrestrial links provided by NIC NET. Internet access is provided by three major network service providers in the country viz., NIC NET, ERNET and VSNL. The BTIS net permits remote login, file transfer, E-mail, etc, as well as connecting to various international networks which are providing up-to-date information support on all aspects in biotechnology ranging from bibliographic information, to sequence analysis and management information.

The network approach had been very useful in the successful implementation of the project as it is enabled to establish a link between diverse groups of scientists working in various interdisciplinary areas of biotechnology. The network encouraged sharing of knowledge and greater interaction amongst the scientific community irrespective of their respective geographical location in the country.

European Molecular Biology Network In India:

The European Molecular Biology Network (EMB net), is a science- based group of 37 Collaborating nodes throughout Europe and a number of nodes outside Europe. EMB net has evolved into a worldwide organisation bringing bioinformatics professionals to work together to serve the expanding fields of genetics and molecular biology. The combined expertise of all nodes allows EMB net to provide Bioinformatics services to the research community and this encompasses more than what a single node can provide bioinformatics at CDFD is recognised at the national node of the EMB net.

The EMB net India Node provides bioinformatics services in the form of browsing bio-molecular sequence databanks, macromolecular structure databanks, genome and other useful databases. It provides in- house services for the comparison and analysis of sequence/ structure/ genome data protein 3-D modelling molecular graphics. The Node also conducts in- house training for the use of commercial software such as GCG and Insight II. The research interests focus on protein sequence and structure analysis, protein modelling, drug design, genome analysis and databases.

An EMB net node was also considered for India and mirrored at Centre for DNA Fingerprinting and Diagnostic (CDFD), Hydrabad. It provides in- house services for the comparison and analysis of sequence/genome data, protein 3-D modelling and molecular graphics. The research interests focus on protein sequence and structure analysis, protein modelling, drug design, genome analysis and databases.

Biotechnology Information System Network In India:

An Apex Centre at the Department of Biotechnology, ten Distributed Information Centres, and 44 Distributed Information sub-centres, established in universities and research institutes of national importance, are fully engaged in this task. All the centres are interlinked through satellite communication system, each providing information support in specific areas of biotechnology and helping in the diffusion of scientific information across the network. Six national facilities have been set up for interactive graphics-based molecular modelling and other biocomputational needs. Four long-term courses at the level of Post M.Sc. Diploma in Bioinformatics, at Poona University Jawaharlal Nehru University, Calcutta University and Madurai Kamaraj University, are fulfilling the long outstanding need for trained human resources in this interdisciplinary area.

The BTIS secretariat working under the department of Biotechnology is a coordinating and developing organisation controlling and guiding the bioinformatics centres.

Need and Significance of the Study:

Now-a-days Information Technology has been changing the Scientific field. The present scenario of the knowledge society, where the communication is in electronic form such as telelearning, teleworking, teleconferencing and video conferencing. Substitute for person-to-person contact must have significant implication for the way which human being satisfies their need for communication with one another.

In recent years several studies has been conducted pertaining to information gathering habits of Scientists, and this study is intended to understand information seeking habits of biotechnology Scientists.

On the basis of this study investigator can identify the problems, which are facing by bio-scientists of biotechnology institutes while gathering needful information and disseminate the information once they received.

Selection of the problems:

The problem for the present study is entitled "Bioinformatics in the Biotechnology Institutes of Northern India: A survey among Scientists". The studies has been conducted to know the bioinformation needs of biotechnology Scientists working in various biotechnology institutes of New Delhi, Ghaziabad, Lucknow and Aligarh.

Definitions:

Bioinformatics:

According to Per Gambeck and Gibas Cyndhia "Bioinformatics is the field of Science in which biology, Computer Science and information technology merge to form a single discipline."

Biotechnology:

According to McGraw-Hill Dictionary of Scientific and Technical Terms "The application of engineering and technological principles to the life sciences."

Institutes:

According to Webster Illustrated Contemporary Dictionary "An established organization or society pledged to some special purpose and work."

India:

According to Oxford English Dictionary "A Country of Southern Asia, lying east of the river Indus and South of Himalayas mountains also called Hindustan."

According to Webster Illustrated Contemporary Dictionary "A republic of the Common Wealth of Nations located on a sub-continental of Asia." Northern India is a part of India.

Survey:

According to Oxford English Dictionary "The examine and ascertain the conditions, situation or value of formally or officially."

Scientists:

According to McGraw-Hill Dictionary of Scientific and Technical Terms "A series of investigations concerned with different problems."

Scope and Limitations of the Study:

The present study is entitled "Bioinformatics in the Biotechnology Institutes of Northern India: A survey among Scientists." The main purpose of the study is to identify, what kind of sources are used by the bio-scientists, and how they generate and disseminate the information once they received. For the study, the variables taken for detailed analysis are bio-scientists working in various institutes of Delhi, Ghaziabad, Lucknow and Aligarh.

Even though hard attempt has been made to make the study as precise and objective as possible certain limitations might have come into the study. The investigator was able to identify some of the major limitations such as:

a) The present study has been conducted only among the bio-scientists working in bio-technology institutes of Northern India (Limitation by class).

b) The geographical area is restricted in Northern India i.e. Delhi, Ghaziabad, Lucknow and Aligarh. (Limitation by Geography).

c) It takes in the account those scientists who have been working or researching in the academic year 2001-2002 (Limitation by Time period).

Objectives of the Study:

The main objectives of the study are:

1. To access the utilization of bio-information by the Scientists of biotechnology Institute of Northern India.
2. To measure the impact of modern technology in the information seeking behaviour of bio-scientists.
3. To know most popular source of information among Scientists.
4. To find out the purpose of using information sources.
5. To identify frequency and purpose of visiting the libraries by bio-scientists of Northern India.
6. To know the problems faced by the bio-scientists in obtaining required information.

7. To identify the methods used by the Scientists to disseminate the information once they received.

Hypotheses:

1. The Scientists are using modern sources of information to meet their information needs. And the CD-ROM Database has an increasing popularity among the Scientists.
2. The Scientists are visiting the library regularly. Most of them use more the primary source of information.
3. The Scientists are very active in participating Conferences and Seminars.
4. The use of Internet and other Computerised Services are play a vital role in the information gathering of Scientists and they are very much familiar with the various types of On-line sources.

Methodology:

The present study is conducted on as sample of 78 bio-scientists of biotechnology institute of Northern India such as, Delhi, Ghaziabad, Lucknow and Aligarh.

There are several techniques are available for data collection such as questionnaire method, personal interviews, telephonic interview, diary method and observation by self etc. Methods like analysis of library records and citation analysis are also used. For this the investigator used questionnaire and observation for data collection.

Method of Data Collection:

For this study the data has been collected through a questionnaire, interview and observation. A questionnaire consisting of 18 question was designed to elicit opinion of the Scientists regarding information seeking behaviour, of them

ORGANISATION OF REPORT:

<u>Chapter-1</u>

Introduction:

An attempt is made to explain the concept of bioinformatics, importants of bioinformatics, Protein Modelling, Genome Mapping, Human Resource Development, Bioinformatics Network, followed by need and Significance of the study, selection of the problems, definitions, objectives of the study, hypothesis and scope and limitations of the study.

<u>Chapter-2</u>

Methodology:

This chapter deals with the selection of the problems, objectives of the study, Hypotheses, methodology, sample population, variable taken, pilot survey, tools and techniques employed used for the study and data analysis.

Chapter-3

Data Collection, Analysis and Interpretation:

This chapter deals with the analysis and interpretation of data collected through questionnaire, interview and observation.

Chapter-4

Conclusion, Findings and Suggestions.

REFERENCES

1. ALTER (O), BROWN (PO) and BOTSTEIN (D). Singular value decomposition for genome-wide expression data processing and modeling. USA, 1997; p.10101-10106.

2. ALTSCHUL (SF), BOGUSKI (W) and WOOTON (JC). Issues in searching molecular sequence databases. *Nature Genetics,* 6, 1994; p.119-129.

3. BAIROCH (A) and BOECKMANN (B). The Swiss Prot Protein sequence databank. *Nucleic Acids Research*, 20 (Supplement), 1992; p.2019-2022.

4. BARKER (WC), GEORGE (DG) and TSUGITA (A). The PIR-International Protein sequence database. *Nucleic Acids Research*, 20 (Supplement), 1992; p.2023-2026.

5. BENDOR (A) and YAKHINI (Z). Clustring gene expression patterns. In RECOMB: Proceedings of the Third Annual International Conference on Computational Molecular Biology, France, 1999; p. 33-42.

6. BENSON (D), LIPMAN (DJ) and OSTELL (J) GenBank. *Nuleic Acids Research*, 34, 1993; p.2963-2965.

7. BOGUSKI (MS). Bioinformatics. Current Opinion in Genetics and Development, 4, 1994; p.383-388.

8. BUNTROCK (RE). Sequence databases: What's in it for me? *Database*, 1991; p.107-109.

9. JAMBECK (Per) and CYNDHIA (Gibas). Developing Bioinformatics, Mumbai; Computer skills, 2001,

10. LANDAU (Sidney I) and BOGUS (Ronald J). Webster Illustrated Contemporary Dictionary: encyclopaedic edition, Chicago; Ferguson, 1978.

11. PARKAR (sybil P). McGraw- Hill Dictionary of Scientific and Technical Terms. 3rd ed., New York; McGraw-Hill, 1984.

12. SIMPSON (JA) and WEINER (ESC). Oxford English Dictionary. Oxford; Clareden, 1989.

CHAPTER-II

METHODOLOGY

- Selection of the Problem
- Objectives of the Study
- Hypotheses
- Methodology
- Sample Population
- Pilot Survey
- Variable taken
- Tools and Techniques Employed
- Data Analysis Method

METHODOLOGY

This Chapter deals with the methodology used in the study and has been discussed under the following headings:

- Selection of the Problem
- Objectives of the Study
- Hypotheses
- Methodology
- Sample Population
- Pilot Survey
- Variables Taken
- Tools and Techniques Employed
- Data Analysis Method

Selection of the Problems:

The problem for the present study is entitled “Bioinformatics in the Biotechnology Institutes of Northern India: A survey among Scientists.”

Objectives of the Study:

The main objectives of the study are:

1. To assess the utilization of bio-information by the Scientists of biotechnology institutes of Northern India.
2. To measure the impact of modern technology in the information seeking behaviour of bio-scientists.
3. To know most popular source of information among Scientists.
4. To find out the purpose of using information sources.
5. To identify frequency and purpose of visiting the libraries by bio-scientists of Northern India.
6. To know the problems faced by the bio-scientists to obtaining required information.
7. To identify the methods used by the scientists to disseminate the information once they received.

HYPOTHESES

1. The Scientists are using modern sources of information to meet their information needs. And the CD-ROM Database has an increasing popularity among the Scientists.
2. The Scientists are visiting the library regularly. Most of them use more the primary source of information.

3. The Scientists are very active in participating Conferences and Seminars.

4. The use of Internet and other Computerised Services are play a vital role in the information gathering of Scientists and they are very much familiar with the various types of on-line sources.

Methodology:

There are several techniques are available for data collection such as questionnaire method, personal interview, telephonic interview, diary method and observation by self etc. Methods like analysis of library records and citation analysis are also used. For this study the investigator used questionnaire interview and observation for data collection.

Questionnaire Method:

Questionnaire method is a useful tool for collecting information from a geographically scattered sample or population groups. This method consists of a careful translation of the objectives of survey into a set of questions, Which may ask for the opinion of the scientis or factual information. The questions are framed in such a way that the answers can be given by checking Yes or No option by selecting one of the possible answers provided in the questionnaire.

Observation:

Observation is the method of acquiring knowledge about the world around us. Observation means systematic viewing of the phenomenon, it is perception with a purpose. It is the oldest and well established technique for collection of data. In measuring, testing characterising human beings, the researcher usually begins with the observable behaviour and historical investigations.

Sample Population:

The collection of large quantity of data from the entire population of Scientists is too large to be adequately covered in a single study. Therefore the total number of 100 questionnaires were administered among Scientists in different Biotechnology Institutes of Northern India during the academic year 2001-2002. A total number of 78 (78.00%) filled questionnaires were returned from the Scientists.

Pilot Survey:

A study preceding the main study usually to check the validity of the study design is known as pilot study or survey. To deciding the present questionnaire or questions were relevant for the purpose of the study, the investigator were distributed questionnaire among 20 scientists of the Bioinformatics in the Biotechnology Institutes of Northern India for the pilot study which was helpful in modifying the questionnaire suitably.

Variable Taken:

In order to achieve the objective of the study following variable is taken for detailed analysis Bio-scientists working in northern India.

Tools and Techniques Employed:

Questionnaire prepared by the investigator supplemented by observation are used as tools for a data collection.

Data Analysis Method:

The data collected through questionnaire are organised and tabulated by using statistical methods, tables and percentages, data collected through observation are used for the formulation of the findings.

CHAPTER-III

DATA COLLECTION, ANALYSIS AND INTERPRETATION

DATA COLLECTION, ANALYSIS AND INTERPRETATION

The problem taken for the present study is "Bioinformatics in the Biotechnology Institutes of Northern India: A survey among the Scientists". The collected data are organised and tabulated by using statistical methods, Tables and percentages. This chapter deals with the analysis and interpretation of the data collected through questionnaire and observation. A total number of 100 questionnaire were distributed among scientists in the Biotechnology Institutes of Northern India, during the academic year 2001-2002, 78 filled questionnaire were returned back from biotechnology scientists.

Use of Information Sources:

The Table-1 indicates the rank order of information sources used by the scientists which reveals the Internet, books, and periodicals are the most used sources, 60(76.92) out of 78 respondents give first rank to the Internet and 7(8.97) give second rank, whereas an equal number of 3(3.85) give 3rd and 4th

Table-1: Use of Information Sources

Sl.No	Information sources	I	II	III	IV	V	Total
1.	Seminars/ Conferences	1 (1.28)	1 (1.28)	9 (11.54)	38 (48.72)	29 (37.18)	78
2.	Information Centres	1 (1.28)	3 (3.85)	9 (11.54)	25 (32.05)	40 (51.28)	78
3.	*Institute* Library	3 (3.85)	17 (21.79)	45 (57.69)	10 (12.82)	3 (3.85)	78
4.	Books/ Periodicals	14 (17.95)	50 (64.10)	11 (14.10)	2 (2.56)	1 (1.28)	78
5.	Internet	60 (76.92)	7 (8.97)	3 (3.85)	3 (3.85)	5 (6.41)	78

(Figures Within Parenthesis Indicate Percentage).
Figures in Roman Letters Indicate Rank Order.

and 5(6.41) give 5th rank. Similarly in the case of books and periodicals 14(17.95) out of 78 give first rank and 50 (64.10) rank as second source of information, whereas 11 (14.10), 2 (2.56) and 1 (1.28) respondents give 3rd, 4th and 5th rank respectively. 3(3.85) give first rank to institute library as the main source of information, whereas 17 (21.78), 45 (57.69), 10 (12.82) and 3 (3.85) respondents give 2nd, 3rd, 4th and 5th rank respectively. It is also evident that the rank position goes to Seminars/ Conferences and Information Centres, as the first rank by 1.28 respondents. 1(1.28) and 3 (3.85) respondents ranked them as second source of information, while both gets third rank from 11.54 respondents each, whereas 38 (48.72), 29(37.18), 25 (32.05) and 40 (51.28) respondents give 4th and 5th rank respectively, the results indicates that Seminars/ Conferences are the less used sources by the scientists, which gets the poorest position.

Use of CD-ROM Databases:

Table-2 shows that the MEDLINE database has by chosen by most of the Scientists i.e. 35 (44.87). 20 (25.64) use GenBank, 10 (12.82) use EMBL (Abridge), 8 (10.26) use PIR-Protein and 5 (6.41) use DNA star. There databases are very

Table-2: Use of CD-ROM databases

Sl. No.	Databases	Scientists
1	MEDLINE	35 (44.87)
2	Gen Bank	20 (25.64)
3	EMBL (Abridge)	10 (12.82)
4	PIR – Protein	8 (10.26)
5	DNA Star	5 (6.41)

(Figures within Parenthesis indicate percentage.)

important for Scientists because they provide current information which help in research work.

Source of Getting Databases:

Table-3 indicates that the scientists get their journals through various sources. 34 (43.59) get through database venders, 30 (38.46) get through Agencies and 14 (17.95) get directly from the producers. The Table reveals that most of the scientists are depending venders to get their volumes and very few approach the publishers directly.

Methods to Access the Journals:

The various sources of getting information were sought under 4th rank. The results are presented in Table-4. Which shows the institute library and the personal copy through the membership were the most common method of having their journals. 3(3.85) put the first rank for personal copy subscription 50(64.10) give first preference for institute library, 20(25.64) for personal copy through membership and 5(6.41) give the first rank for catalogue. 57(73.08) give second preference. 15(19.23) get their journals from institute library, 4(5.13) personal copy through membership and 2(2.56) personal copy through subscription.

Table-3: Source of getting databases

Sl. No.	Sources	No. of Scientists
1	Database Venders	34 (43.59)
2	Agencies	30 (38.46)
3	Directly from Producers	14 (17.95)

(Figures Within Parenthesis Indicate Percentage.)

Table-4: Methods to Access the Journals

Sl.No	Sources	I	II	III	IV	Total
1.	Personal Copy Through Subscription	3(3.85)	2(2.56)	4(5.13)	69(88.46)	78
2.	Personal Copy Through Membership	20(25.64)	4(5.13)	52(66.67)	2(2.56)	78
3.	From Catalogues	5(6.41)	57(73.08)	12(15.38)	4(5.13)	78
4.	Institute Library	50(64.10)	15(19.23)	10(12.82)	3(3.85)	78

(Figures Within Parenthesis Indicate Percentage.)

52 scientists (66/67) marked the 3^{rd} rank for personal copy acquired by membership, 12 respondents (15.30) gave the same rank to catalogue. 10(12.82) for Institute library and 4(5.13) for personal copy through subscription. Among the scientists 49(88.46) give 4^{th} rank for personal copy acquired by subscription, while 2(2.56) for personal copy through membership, 4(5.13) for catalogue as a tool and 3(3.85) for Institute Libraries.

Purpose of Using Periodicals:

The Table-5 depicts the purpose of using periodicals among scientists that 35(44.87) respondents use for research purpose, 30 (38.46) respondents use for updating knowledge, 7 (8.97) respondents use for general awareness and 6 (7.69) respondents use for teaching purpose out of 78 respondents from the scientists.

Purpose of Visiting the Library:

In order to assess the frequency of visit to the library, the gap has been classified into 5 different categories as shown in Table-6. The analysis declared that 28(39.89) respondents visit

Table-5: Purpose of Using Periodicals

Sl. No	Purpose	No. of Scientists
1.	Updating Knowledge	30(38.46)
2.	Research	35(44.87)
3.	Teaching Work	6(7.69)
4.	General Awareness	7(8.97)

(Figures Within Parenthesis Indicate Percentage.)

the library to get books issued followed by 24 (30.77) respondents frequently, 19(24.36) respondents sometimes, 4(5.13) respondents rarely and 3(3.85) are not at all get issued the books from the library. 9(11.54) scientists visit the library to use periodicals usually, followed by 18(23.08) respondents told that they never use periodicals. 22(28.20) respondents from the scientists visit the library to use the reference materials very frequently, 26 (33.33) respondents used frequently, 24(30.77) respondents use sometimes, 4(5.13) respondents use rarely and 2(2.56) respondents never use. 18(23.08) scientists visit the library usually to read the Newspapers/ Magazines. 23(29.49) visit frequently, 4(5.38) rarely and 5(6.41) stated that they never visit the library for reading the newspapers/ magazines.

The Table shows that only 2(2.56) visit their library very frequently to use audio visual materials. Among the scientists 4(5.13) frequently, 6(7.60) sometimes, 7(8.97) rarely visit the library for the same purpose. And it is notable that a major 59(75.64) scientists never visit the library to use the audio materials.

Internet is an essential Information Sources, especially in the field of Science and Technology which will provide the fresh valuable information to the scientists. Table-6 discloses that

Table-6: Purpose of Visiting the Library by Scientists

Sl.No	Purpose	VF	F	S	R	N	Total
1.	To get books issued	28 (39.89)	24 (30.77)	19 (24.36)	4 (5.13)	3(3.85)	78
2.	To use Periodicals	9 (11.54)	18 (23.08)	35 (44.87)	14 (17.95)	2(2.56)	78
3.	To use reference materials	22 (28.20)	26 (33.33)	24 (30.77)	4 (5.13)	2(2.56)	78
4.	To read newspapers/ magazines	18 (23.08)	23 (29.49)	20 (25.64)	12 (15.38)	5(6.41)	78
5.	To use audiovisual materials	2 (2.56)	4 (5.13)	6 (7.69)	7(8.97)	59(75.64)	78
6.	To browse internet	17 (21.79)	6 (7.69)	8 (10.26)	4(5.13)	43(55.13)	78

VF = Very Frequently, F = Frequently, S = Sometimes, R = Rarely, N = Never

(Figures Within Parenthesis Indicate Percentage.)

17(21.79) scientists visit their library to browse the Internet usually. 6(7.69) visit frequently, 8(10.26) sometimes, 4(5.13) only rarely, and it is notable that 43(55.13) never visit the library to use Internet.

The Table reveals that the scientists are visiting the library irregularly as their department libraries are sufficient to satisfy their needs.

Use of the Library Services:

The Table-7 indicates the rank order of services used by the scientists that the reference and circulation are the most used services, 52(16.67) out of 78 respondents marked first rank to the reference and 13(16.67) gave second rank, whereas 7(8.97), 3 (3.8), 1 (1.28), 1 (1.28) and 1 (1.28) respondents gave 3rd, 4th, 5th, 6th and 7th ranks respectively. Similarly in the case of circulation 13 (16.67) out of 78 give first rank and 18 (23.08) give second rank, whereas 10 (12.82) give 3rd, 4th, 5th, 6th and 7th rank respectively. Photocopying service used by 8 (10.26) out of 78 respondents. The scientists 8(10.26) give first rank, whereas 20(25.64), 20(25.64), 13 (16.67), 9 (11.54), 5 (6.41) and 3 (3.85) give 2nd, 3rd, 4th, 5th, 6th and 7th rank respectively for Photocopy Service. Selective Dissemination of Information (SDI) is used by,

Table-7: Use of the Library Services

Sl.No	Services	Scientists							
		I	II	III	IV	V	VI	VII	Total
1.	Circulation	13 (16.67)	18 (23.08)	10 (12.82)	8 (10.26)	8 (10.26)	8 (10.26)	13 (16.67)	78
2.	Reference	52 (66.67)	17 (16.67)	7 (8.97)	3 (3.85)	1 (1.28)	1 (1.28)	1 (1.28)	78
3.	Indexing	1 (1.28)	7 (8.97)	13 (16.67)	16 (20.51)	14 (17.95)	18 (23.08)	9 (11.54)	78
4.	Abstracting	2 (2.56)	12 (15.38)	10 (12.82)	7 (8.97)	24 (30.77)	17 (21.79)	8 (10.26)	78
5.	Photocopying	8 (10.26)	20 (25.64)	20 (25.64)	13 (16.67)	9 (11.54)	5 (6.41)	3 (3.85)	78
6.	CAS	2 (2.56)	1 (1.28)	16 (20.51)	24 (30.77)	9 (11.54)	16 (20.51)	10 (12.82)	78
7.	SDI	3 (3.85)	5 (6.41)	5 (6.41)	8 (10.26)	10 (12.82)	13 (16.67)	34 (43.59)	78

(Figures Within Parenthesis Indicate Percentage).

Figures in Roman Letters Indicate Rank Order

3 (3.85) out of 78 scientists who gives first preference whereas 5 (6.41), 5 (6.41), 8 (10.26), 10(12.82), 13 (16.67) and 34 (43.59) give 2nd, 3rd, 4th, 5th, 6th and 7th ranks respectively. It is also evident that then position goes to abstracting and Current Awareness Service (CAS), both gets first rank from 2.56 respondents each, 12(15.38) and 1(1.28) give second rank, whereas 10(12.82), 7(8.97), 24(30.77), 17(21.79), 8(10.26) and 16(20.51), 24(30.77), 9(11.54), 16(20.51), 10(12.82) respondents give 3rd, 4th, 5th, 6th and 7th ranks respectively. 1(1.28) respondent give first rank to indexing, whereas 7 (8.97), 13 (16.67), 16 (20.51), 14 (17.95), 18 (23.08) and 9 (11.54) respondents give 2nd, 3rd, 4th, 5th, 6th and 7th rank respectively, the result also indicates that indexing is the less used service among the respondents.

Purpose of Seeking Information:

The Table-8 indicates the rank order of information seeking purpose of the scientists. The scientists seek information mainly for the career development and to solve immediate practical problems. 60 (76.92) out of 78 respondents give first rank to the career development and 11 (14.1) give second rank, whereas 4 (5.13), 1 (1.28) and 2 (2.56) respondent give 3rd, 4th, and 5th rank respectively. Similarly to solve immediate practical problems, 17

Table-8: Purpose of Seeking Information

Sl.No.	Purposes	Scientists					
		I	II	III	IV	V	Total
1.	For Career Development	60 (76.92)	11 (14.10)	4 (5.13)	1 (1.28)	2 (2.56)	78
2.	To Solve Immediate Practical Problems	17 (21.79)	33 (42.31)	9 (11.54)	13 (16.67)	6 (7.69)	78
3.	Guide the Students	2 (2.56)	11 (14.10)	19 (24.36)	27 (29.49)	19 (24.36)	78
4.	Personal Ego and Prestige	1 (1.28)	12 (15.38)	23 (29.49)	21 (26.92)	21 (26.92)	78
5.	To Write Research Paper	4 (5.13)	13 (16.67)	20 (25.64)	12 (15.38)	29 (37.18)	78

(Figures Within Parenthesis indicate percentage).

Figures in Roman Letters Indicate Rank Order.

(21.79) out of 78 give first rank and 33 (42.31) out of 78 give second rank, whereas 9 (11.54), 13 (16.67) and 6 (7.69) respondents give 3rd, 4th and 5th rank respectively. 4 (5.13) respondents give first rank to write research paper, whereas 13 (16.67), 20 (25.64), 20 (25.64), 12 (15.38) and 29 (37.18) respondents give 2nd, 3rd, 4th and 5th rank for the research paper writing purpose respectively. 2 (2.56) respondents give first rank to guide the students, whereas 11 (14.10), 19 (24.36), 27 (29.49) and 19 (24.36) respondents give 2nd, 3rd, 4th and 5th rank respectively. A single respondent 1(1.28) give first rank to personal ego and prestige, whereas 12 (15.38), 23 (29.49), 21 (26.92) and 21 (26.92) respondents give the 2nd, 3rd, 4th and 5th rank respectively. The result also indicates that personal ego and prestige are the minor reason for seeking information among the respondents.

Use of Information Sources:

The Table 9 indicates the rank order of information sources used by the Scientists which reveals the Internet, books and periodicals are the most used source, 54 (69.23) out of 78 respondents give first rank to the Internet and (2.56) give 2nd rank, whereas 4 (5.13) give 3rd and 4th rank, 9 (11.54), 2 (2.56)

Table-9: Use of information sources

Sl. No.	Information Sources	I	II	III	IV	V	VI	VII	VIII	IX	Total
1.	Books	12 (15.38)	18 (23.08)	29 (37.18)	9 (11.54)	1 (1.28)	1 (1.28)	2 (2.56)	2 (2.56)	4 (5.13)	78
2.	Periodicals	6 (7.69)	8 (10.26)	10 (12.82)	10 (12.82)	14 (17.95)	12 (15.38)	4 (5.13)	8 (10.26)	6 (7.69)	78
3.	News papers/ Press Clippings	1 (1.28)	5 (6.41)	9 (11.54)	13 (16.67)	10 (12.82)	15 (19.23)	15 (19.23)	9 (11.54)	1 (1.28)	78
4.	Government document	3 (3.85)	1 (1.28)	1 (1.28)	9 (11.54)	7 (8.97)	15 (19.23)	16 (20.51)	15 (19.23)	11 (14.10)	78
5.	Dissertation/ Thesis	1 (1.28)	5 (6.41)	3 (3.85)	11 (14.10)	19 (24.36)	12 (15.38)	10 (12.82)	14 (17.95)	3 (3.85)	78
6.	Conference/ Seminar Proceedings	3 (3.85)	2 (2.56)	13 (16.67)	14 (17.95)	19 (24.36)	13 (16.67)	7 (8.97)	3 (3.85)	4 (5.13)	78
7.	CD- ROM database	4 (5.13)	5 (6.41)	8 (10.26)	6 (7.69)	3 (3.85)	6 (7.69)	9 (11.54)	8 (10.26)	29 (37.18)	78
8.	On-line database	2 (2.56)	37 (47.44)	4 (5.13)	4 (5.13)	7 (8.97)	2 (2.56)	4 (5.13)	15 (19.23)	3 (3.85)	78
9.	Internet	54 (69.23)	2 (2.56)	4 (5.13)	4 (5.13)	1 (1.28)	1 (1.28)	9 (11.54)	2 (2.56)	1 (1.28)	78

(Figure Within Parenthesis Indicate Percentage).

Figures in Roman letters indicate rank order.

and 1 (1.28) each give 5^{th} and 6^{th} rank, 9 (11.54), 2 (2.56) and 1 (1.28) give 7^{th}, 8^{th} and 9^{th} ranks respectively. Similarly in he case of books 12 (15.38) out of 78 give first rank and 18 (23.08) rank as second source of information, whereas 29 (37.18) and 9 (11.54) give 3^{rd} and 4^{th} rank respectively. 1 (1.28) each give 5^{th} and 6^{th} rank, 2 (2.56) each give 7^{th} and 8^{th} rank to periodicals as the main source of information and 8 (10.26) give 2^{nd} rank, whereas 10 (12.82) each give 3^{rd} and 4^{th} rank, 14 (17.95), 12 (15.38), 4 (5.13), 8 (10.26) and 6 (7.69) respondents give 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. It is also evident that the rank position goes to the first from 4 (5.13) respondents to CD-ROM database and 5 (6.41) give second rank, whereas 8 (10.26), 6 (7.69), 3 (3.85), 6 (7.69), 9 (11.54), 8 (10.26) and 29 (37.18) respondents give 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. Similarly in the case of Government documents and conference/ seminar proceedings 3 (3.85) out of 78 give 1^{st} rank, 1 (1.28) each rank as 2^{nd} and 3^{rd} source of information, where as 9 (11.54), 7 (8.97), 15 (19.23), 16 (20.51), 15 (19.23) and 11 (14.10) respondents give 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. While conference/ seminar proceedings are treated by (2.56), 13 (16.67), 14 (17.95), 19 (24.36), 13 (16.67), 7 (8.97), 3 (3.85) and 4 (5.13) respondents by giving 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. Similarly in the case of newspapers/ press clippings and dissertation/ theses out of 78 give first rank and 5(6.41) ranks second source of information, where as 9(11.54), 13(16.67), 10(12.82), 15(19.23), 9(11.54) and 1(1.28) respondents give3rd, 4^{th}, 5^{th}, 6^{th} and 7^{th}, 8^{th} and 9^{th} ranks respectively. While for dissertation/ theses, 3(3.85), 11(14.10), 19(24.36), 12(15.38), 10(12.82), 14(17.95) and 3(3.95) respondents give 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. 2(2.56) respondent give first rank for On-line database and 37(47.44), 4(5.13), 4(5.13), 7(8.97), 2(2.56), 4(5.13), 15(19.23) and 3(3.85) respondents give 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 6^{th}, 7^{th}, 8^{th} and 9^{th} ranks respectively. The result also indicates that newspapers/ press clippings and dissertation/ theses are the least used reason of information among the respondents.

Use of Internet Facility:

The Table 10 indicates the use Internet facility among the scientists i.e. 35(44.87) respondents use E-mail (Electronic-mail). 5(6.41) respondents use bulletin board system, 4 (5.13) use file transfer protocol, 3(3.85) use Tell net, 18 (23.08) use net, 6(7.69) respondent use WWW (World Wide Web), 4(5.13) use document management, 2(2.56) use E-journal (Electronic-Journal) and 1(1.28) use the Chatmode.

Table-10: Use of Internet Facility

Sl.No	Services	No. of Scientists
1.	E-mail (Electronic-mail)	35 (44.87)
2.	Bulletin Board System	5 (6.41)
3.	File Transfer Protocol	4 (5.13)
4.	Tell Net	3 (3.85)
5.	Use Net	18 (23.08)
6.	WWW (World Wide Web)	6 (7.69)
7.	Document Management	4 (5.13)
8.	Electronic Journal	2 (2.56)
9.	Chatmode	1 (1.28)

(Figures Within Parenthesis Indicate Percentage).

Now-a-days Internet is the main stream information transfer channel which has an important role in the field of Science and Technology. It provides current information to the Scientists. Within minutes at anywhere in the world.

Awareness of on-Line Services:

The data shown in Table-11 indicate that 30(38.46) of respondents are only aware of the on-line services, 30(38.46) use and 18(23.08) are aware and use the E-mail service.

In case of Internet the Table reveals that 35(44.87) of respondents are only aware about it. 35(44.87) use and 8(10.26) are aware and using the Internet Service.

The investigators observation reveals that many of the scientists use the E-mail and Internet facility for their personal purpose.

Use of Online-Database:

The Table-12 indicates that 75(96.15) of scientists use BLAST & FASTA on-line database and only 3(3.85) make use of US Patent On-line Database.

Table-11: Awareness of On-line Services

SI.No.	Services	Aware Only	Use Only	Aware & Use	Total
1.	E-mail	30 (38.46)	30 (38.46)	18 (23.08)	78
2.	Internet	35 (44.87)	35 (44.87)	8 (10.26)	78

(Figures Within Parenthesis Indicate Percentage).

Table-12: Use of On-line Databases

SI.No.	On-line Databases	No. of Scientists
1.	BLAST & FASTA	75 (96.15)
2.	US Patent	3 (3.85)

(Figures Within Parenthesis Indicate Percentage.)

Channels of Information Dissemination:

It is common knowledge that dissemination of information is an important in today's context. The Biotechnology Institutes have very magnificent role in the process of dissemination of information. Respondents were asked to state how they disseminate information once they got. The responses were as follows: lecturing, publishing, seminar presentation, research reports, radio and television broadcasting, telephone and newspapers. The data given in Table 13 show that the research reports were chosen by the most of the scientists 22(28.21). 20(25.64) use lecturing, 16(20.51) use publishing, 14(17.95) use seminar presentation, 3(3.85) use newspapers and 1(1.28) use telephone for the same purpose.

Table-13: Channels of Information Dissemination

Sl. No.	Channels	No of Scientists
1.	Lecturing	20 (25.64)
2.	Publishing	16 (20.51)
3.	Seminar Presentation	14 (17.95)
4.	Research Reports	22 (28.21)
5.	Radio	1 (1.28)
6.	Television	1 (1.28)
7.	Telephone	1 (1.28)
8.	Newspapers	3 (3.85)

(Figures Within Parenthesis Indicate Percentage)

CHAPTER-IV

CONCLUSION, FINDINGS AND SUGGESTIONS

CONCLUSION, FINDINGS AND SUGGESTIONS

CONCLUSION:

This study is sought to examine the Bioinformatics in the Biotechnology Institutes of Northern India: A survey among Scientists". The studies has been conducted to know the bioinformation needs of biotechnology scientists of New Delhi, Ghaziabad, Lucknow and Aligarh

Most of the objectives are meet satisfactorily and most of the bio-scientists of Biotechnology Institutes think that on-line service are essential in the Biotechnology Institutes but ratio of the using online E-journals on the Internet is different because it is dependent upon the bio-Scientists need and interest.

On-line technology has an impact on all most all aspects of Biotechnology Institutes and information work. Biotechnology Institutes in the developing world appreciate the importance of the information systems of the developed world.

Currently information and associated communications are gradually increasing their impact on our society and the values that we espouse and it is important to note some developing issues.

The use of Internet and other computerised services are play a vital role in the information gathering of bio-Scientists and they are very much familiar with the various types of on-line sources.

FINDINGS:

- Internet is the most used information sources among the bio-technology scientists followed by books and periodicals (Table-1).
- Seminars/Conference and Information Centres are the minimum used information sources. (Table-1).
- MEDLINE is the most used and DNA star is the less used CD-ROM database services among the Scientists (Table-2).
- Gen Bank and EMBL (Abridge) also have an increasing popularly among Scientists (Table-2).
- A 43.59 percent of Scientists are depending the database venders to have their databases (Table-3).

- Most of the Scientists (64.10) are depends on Institute Library to read and only a few Scientists (3.85) subscribe Journals personally (Table-4).
- A high percentage of scientists using journals for their research purpose.
- A good percentage of Scientists (38.46%) read the Journals to update their knowledge. (Table-5).
- Audio-Visual Collections of the Library has not earned much popularity even today. (Table-6).
- Reference Section is one of the most attractive Service of Library to the Scientists. (Table-6) (Table-7).
- Newspapers/Magazines are also used by a good numbers of Scientists. (Table-6).
- Only a few percent (21.79%) Scientists are visiting the Library to browse the Internet (Table-6).
- SDI and CAS Services are very rarely used by the Scientists (Table-7).
- A major portion of Scientists (76.92%) seek the Information for their Career Development. (Table-8).

- To solve the immediate practical problems is also one of the main reason for seeking Information among Scientists (21.79) (Table-8).
- Only a few percentage of Scientists (5.13) seek Information for writing research papers. (Table-8).
- 44.87 percentage of scientists use the internet for mailing purpose (Table-10).
- Use net and WWW (World Wide Web) are the Services which is used more by the scientists for the research purpose. (Table-10).
- Internet chat mode has not earned much popular among the scientists (Table-10).
- All the Scientists are aware of On-line Services. (Table-11).
- Among the On-line Databases BLAST & FASTA is the most popular database among bio-technology scientists (Table-12).
- US Patent is used by only a 3.85 percent of Scientists (Table-12).
- Scientists prefer research reports as a channel to disseminate the information once they received followed by lecturing and seminar presentation (Table-13).

- Very few Scientists are using the Mass media, such as Radio, Television and Telephone for the dissemination of Scientific Information (Table-13).

SUGGESTIONS:

- The libraries may give more consideration for the reference section as it is the most used services by the Scientists
- Library as the center of knowledge and Information, the Internet facility should be in the Institute Library is to provide latest information.
- The Scientists should have been provided the proper guidance and orientation in using the Internet and other collections and services.
- Specialised Services of Institute Library such as SDI & CAS are not properly utilized by the Scientists. The Libraries and Information Centres have to conduct proper orientation and awareness programmes on CAS/SDI Services among the Scientists.
- Online databases, as they are getting more popularity, have to be capable to satisfy the Scientists with the feasibility and possibility.

APPENDICES

i) Questionnaire Administered to Scientists

ii) List of Bioinformatics Centres Visited

iii) A List of Bioinformatics Centres in India

BIOINFORMATICS IN THE BIOTECHNOLOGY INSTITUTES OF NORTHERN INDIA: A SURVEY AMONG SCIENTISTS

"QUESTIONNAIRE ADMINISTERED TO SCIENTISTS"

To,

Sir/ Madam,

I am conducting a survey on the above topic for the award of **M.L & I. Sc.** degree from Aligarh Muslim University. My dissertation supervisor is ***Mr. NAUSHAD ALI P.M.*** (Lecturer), Department of Library and Information Science, A.M.U., Aligarh.

With a view to collect accurate, comprehensive primary data required for the study I have designed the enclosed questionnaire and I am seeking your kind co-operation in filling up the same. The data supplied by you will be immense value for this study and will be used for the dissertation purpose only.

I am aware of your busy schedule but without your help and assistance my study will be incomplete. I shall be grateful if you would kindly spare some time from your busy schedule and fill up the Questionnaire.

Thanking you and with warm regards.

Yours faithfully,

Irshad Ahmad

(IRSHAD AHMAD)

APPENDIX-I

QUESTIONNAIRE

Please fill in the information in the blank space or put a tick mark or give the code number as applicable in each case.

General Information

Name	:	....
Name of Institution/ Organisation	:	....
Designation	:	....
Academic Qualification	:	....
Topic of Research (if any)	:	....

1. From where do you normally get the information whenever you need them?

		Rank order
(i)	Seminars/ Conferences	[ ]
(ii)	Information centres	[ ]
(iii)	Institute Library	[ ]
(iv)	Books/ Periodicals	[ ]
(v)	Internet	[ ]
(vi)	Any other (please specify)	

2. Do you use CD-ROM database regularly? [Yes/ No]

   (i) If yes, please mention [ ]

(a) MEDLINE [ ]

(b) Gen Bank [ ]

(c) EMBL (Abridged) [ ]

(d) PIR – Protein [ ]

(e) DNA star [ ]

(ii) If no please mention the reason..................................

3. How do you collect databases?

(i) Database venders [ ]

(ii) Agencies [ ]

(iii) Directly from producers [ ]

(iv) Any other (please specify) [ ]

4. Do the current database satisfy your needs? [Yes/No]

5. Do you think databases are more useful than any other sources in your field? [Yes/ No]

6. Do you read journals regularly? [Yes/ No]

(A) How do you get access to the journals? Rank order

(i) Personal copy through subscription [ ]

(ii) Personal copy through membership [ ]

(iii) From catalogues [ ]

(iv) Institute library [ ]

(v) Any other (please specify)..

(B) Please indicate the purpose of using periodicals.

(a) Updating knowledge [ ]

(b) Research [ ]

(c) Teaching work [ ]

(d) General awareness [ ]

(e) Any other (please specify)...

7. Do you visit the library?

		VF	F	S	R	N
(i)	To get books issued	[ ]	[ ]	[ ]	[ ]	[ ]
(ii)	To use periodicals	[ ]	[ ]	[ ]	[ ]	[ ]
(iii)	To use reference materials	[ ]	[ ]	[ ]	[ ]	[ ]
(iv)	To read newspapers/ magazines	[ ]	[ ]	[ ]	[ ]	[ ]
(v)	To use audiovisual materials	[ ]	[ ]	[ ]	[ ]	[ ]
(vi)	To browse internet	[ ]	[ ]	[ ]	[ ]	[ ]

VF=Very Frequently, F=Frequently, S=Sometimes, R=Rarely, N=Never

8. Which of the library services listed below do you use regularly?

		Rank order
(i)	Circulation service	[ ]
(ii)	Reference	[ ]
(iii)	Indexing service	[ ]
(iv)	Abstracting service	[ ]
(v)	Photocopying service	[ ]
(vi)	Current awareness service	[ ]

(vii) SDI service [ ]

9. Why do you seek information?

Rank order

(i) For career development [ ]

(ii) To solve immediate practical problems [ ]

(iii) To guide students [ ]

(iv) Personal ego and prestige [ ]

(v) To write research papers [ ]

10. What are the difficulties which you have faced to obtaining required information and keeping update with the advances in your field?

(i) Lack of time [ ]

(ii) Inadequate library resources [ ]

(iii) Information scattered in too many sources [ ]

(iv) Lack of access library material due to Library rules/ procedure [ ]

(v) Lack of suitable abstracting journals/ services [ ]

(vi) Interdisciplinary nature of literature [ ]

(vii) Any other (please specify)..

11. Do you attend conferences, seminars, workshop, lectures related to your subject?

(i)	Very frequently	[ ]
(ii)	Frequently	[ ]
(iii)	Sometimes	[ ]
(iv)	Rarely	[ ]
(v)	Never	[ ]

12. Please indicate the usefulness of the following sources of information according to their importance.

		Rank order
(i)	Books	[ ]
(ii)	Periodicals	[ ]
(iii)	Newspapers/ Press clippings	[ ]
(iv)	Government document	[ ]
(v)	Dissertation / Theses	[ ]
(vi)	Conference/ Seminar proceedings	[ ]
(vii)	CD-ROM database	[ ]
(viii)	On-line database	[ ]
(ix)	Internet	
(x)	Any other (please specify)..	

13. Do you use computerised information services? [Yes/No]

If Yes

(A) Which type of computerised service you are getting?

(i) Literature search within the library [ ]

(ii) Literature search through local network [ ]

(iii) Literature search through national network [ ]

(iv) Literature search through international network [ ]

14. Do you use the following internet services? [Yes/ No]

		Yes	No
(i)	E-mail (Electronic mail)	[ ]	[ ]
(ii)	Bulletin board system	[ ]	[ ]
(iii)	File Transfer Protocol	[ ]	[ ]
(iv)	Tel net	[ ]	[ ]
(v)	Use net	[ ]	[ ]
(vi)	WWW (World wide web)	[ ]	[ ]
(vii)	Document management	[ ]	[ ]
(viii)	Electronic journal	[ ]	[ ]
(ix)	Chatmode	[ ]	[ ]

(x) Any other (please specify)..

15. Do you use on-line data bases? [Yes/ No]

16. Show your familiarity with the following on-line services.

		A	NA	U	NU
(i)	E-mail	[ ]	[ ]	[ ]	[ ]

(ii) Internet [ ] [ ] [ ] [ ]

A=Aware, NA = Not aware, U=Used, NU = Not used.

17. Which on-line database you use mostly?

(i) BLAST & FASTA [ ]

(ii) US Patent [ ]

(iii) Any other (please specify) ....

18. How do you disseminate the information once you received?

Please use [ ] marks to answer the following questions.

(i) Lecturing [ ]

(ii) Publishing [ ]

(iii) Seminar presentation [ ]

(iv) Through research report [ ]

(v) Radio [ ]

(vi) Television [ ]

(vii) Telephone [ ]

(viii) Newspapers [ ]

(ix) Any others (please specify)......................................

SUGGESTIONS..

....

....

....

....

Thanks

APPENDIX-II

LIST OF BIOINFORMATICS CENTRES VISITED

1. Aligarh Muslim University, Aligarh.
2. All India Institute of Medical Sciences, New Delhi.
3. Amity Institute of Biotechnology and Bioinformatics, Noida (U.P.).
4. Biotech Consortium India Limited, New Delhi.
5. Central Drug Research Institute, Lucknow.
6. Central Institute of Medical & Aromatic Plants, Lucknow, (U.P.)
7. Indian Agricultural Research Institute, New Delhi.
8. Indian Institute of Technology, New Delhi.
9. Jai Prakash Institute of Information Technology, Noida (U.P.).
10. Jamia Millia Islamia University, New Delhi.
11. Jawaharlal Nehru University, New Delhi.
12. National Institute of Immunology, New Delhi.
13. University of Delhi (North Campus), New Delhi.
14. University of Delhi (South Campus), New Delhi.
15. Santosh Medical College, Ghaziabad (U.P.).

APPENDIX-III

A LIST OF BIOINFORMATICS CENTRES IN INDIA

1. Aligarh Muslim University, Aligarh.
2. All India Institute of Medical Sciences, New Delhi.
3. Amity Institute of Biotechnology and Bioinformatics, Noida (U.P.).
4. Anna University, Madras.
5. Assam Agricultural University, Jorhat, Assam.
6. Banaras Hindu University, Varanasi.
7. Banasthali Vidyapeeth, Banasthali, Rajasthan.
8. Bharatidasan University, Tiruchirappalli, Tamil Nadu.
9. Biotech Consortium India Limited, New Delhi.
10. Birla Institute of Scientific Research, Jaipur, Rajasthan.
11. Bose Institute, Kolkatta.
12. Central Drug Research Institute, Lucknow.
13. Central Institute of Fresh Water Acquaculture (CIFA), Bhuvaneshwar.
14. Central Institute of Medicinal and Aromatic Plants, Lucknow, (U.P.).

15. Centre for Cellular and Molecular Biology, Hyderabad.

16. Central Sericulture Research and Training Institute, Mysore.

17. Devi Ahliya Vishwavddhyalay, Indore.

18. Dr. YS. Parmar University of Horticulture and Forestry, Solan (H.P.).

19. G.B. Pant University of Agriculture and Technology, Pantnagar (U.P.).

20. Guru Nanak Dev University, Amritsar.

21. Himachal Pradesh University, Shimla.

22. Indian Agricultural Research Institute, New Delhi.

23. Indian Institute of Science, Bangalore.

24. Indian Institute of Technology, Kharagpur, West Bengal.

25. Indian Institute of Technology, New Delhi.

26. Indian Veterinary Research Institute, Izatnagar.

27. Institute of Micorbial Technology, Chandigarh.

28. Jai Prakash Institute of Information Technology, Noida (U.P.).

29. Jamia Millia Islamia University, New Delhi.

30. Jawaharlal Nehru University, New Delhi.

31. Kerala Agriculture University, Thrissur.

32. Mahatama Gandhi Institute of Medical Sciences, Wardha.

33. MS. Swaminathan Research Foundation, Chennai.

34. MS. University of Vadodra, Vadodra.

35. National Dairy Research Institute, Karnal.

36. National Environmental Engineering Research Institute, Nagpur.

37. National Institute of Immunology, New Delhi.

38. National Institute of Oceanography, Goa.

39. North Eastern Hill University, Shillong, Meghalaya.

40. Pondicherry University, Goa.

41. Punjab Agricultural University, Ludhiana, Punjab.

42. Regional Research Laboratory, Jammu & Kashmir.

43. Tamil Nadu Agricultural University, Coimbatore.

44. Tamil Nadu Veterinary and Animal Science University, Madras.

45. Tata Memorial Centre, Mumbai.

46. Tropical Botanical Garden and Research Institute, Tiruvananthapuram, Kerala.

47. University of Kolkatta.

48. University of Delhi (North Campus), New Delhi.

49. University of Delhi (South Campus), New Delhi.

50. University of Poona, Pune.